DOCUMENT RESUME

ED 153 729 PS 009 892

AUTHOR Tucker, Eugene
TITLE The Follow Through Planned Variation Experiment: Threats to Construct Validity.
PUB DATE Mar 78
NOTE 10p.; Paper presented at the Annual Meeting of the American Educational Research Association (Toronto, Canada, March 27-31, 1978); ; Best copy available

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Compensatory Education Programs; Educational Objectives; Elementary Education; *Primary Education; *Problems; Program Development; *Program Evaluation; Test Selection; *Validity
IDENTIFIERS Planned Variation; *Project Follow Through

ABSTRACT

This paper explores problems of construct validity in the evaluation of the Follow Through Planned Variation program. Comments center around political ramifications of the program evaluation design, insufficient knowledge on the part of program planners, the diversity of the experimental options, and the instability of the sample. It is suggested that the political problem of whether the program was to operate in a service or an experimental mode resulted in confusion over what sort of information an evaluation was to provide; this made it impossible to construct and validate instruments for the multidimensional program that Project Follow Through was. It was assumed that expected program impact would be great enough to overcome the marginal biases of the standardized tests used. The fact that planners were forced to start the evaluation with insufficient information about constructs being addressed by different program developers, and insufficient time to develop or select an appropriate testing battery, resulted in a test battery of questionable construct validity. Also, because of the longitudinal nature of the evaluation, evaluators were not free to choose alternate batteries. Another problem resulted from the varying number of replications: among the 22 experimental options funded, replications ranged from one to 14. Finally, sample attrition exceeded 50 percent. It is concluded that more attention to construct validity is needed if meaningful educational evaluation is to be done. (SB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED153729

THE FOLLOW THROUGH PLANNED VARIATION EXPERIMENT: THREATS TO CONSTRUCT VALIDITY

BEST COPY AVAILABLE

By
Eugene Tucker
U. S. Office of Education
400 Maryland Avenue SW
Washington, D.C. 20202

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Eugene Tucker

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Presented at
the annual meeting of the
American Educational Research Association
Toronto Ontario, Canada
March 28, 1978

PS 009892

The comments and opinions expressed in this paper are solely those of the author and must not be construed as the official position of the U.S. Office of Education.

Over the last decade the Follow Through Program has provided a group of bright energetic specialists in education an opportunity to test their theories in a variety of school districts and classrooms throughout the country. Scholars are likely to spend the next decade applying their erudition to the often characterized ill fated heap we sympathetically call the final report on the Follow Through Planned Variation Experiment. The task is monumental given the several thousand pages and the dozen or so volumes that comprise the report. Only a few have dared to attempt a reading of the full treatise. Discussions of the type we are having today may serve to heighten the curiosity of educational researchers and practioners the world over, thus exposing our work to the kind of criticism that will lead to expanded knowledge about educating children from poor families. This was a goal sought in the 1960's when the program was initiated. Let us reflect for a moment on the decade of the 1960's in order that the views to be expressed in this conference can be placed in their proper perspective.

The Follow Through Program was legislated in response to a social crisis. Our national government was faced with ever increasing demands from its varied population for freedom and equality. The social upheavel was unlike any this country had ever witnessed. Many of the best minds in and out of government thought that a solution to the social ills was through education. To this end a massive all encompassing approach was devised. The program would be comprehensive requiring educational innovations, involvement of parents and the community at large, and providing nutritional and medical care. The program was

considered from its inception as service and was expected to be expanded to address the full universe of need. This did not occur and in fact the service notion itself was revised within the U. S. Office of Education. The research mode that followed was not well articualted to participating individuals, institutions, or legislators. This may be the underlying cause for the ultimate difficulties encountered in the evaluation. In the next few pages I will explore some of the threats to construct validity in the Follow Through Planned Variation Experiment.

Construct validation is concerned with the assessment of the degree to which pre-defined traits or attributes have been measured. Underlying this definition is the requirement that the traits or attributes can be measured by several approaches. Further, the logic and use of the validation process must be considered. In education the proposed interpretation of the test must be validated along with the test itself. Cronbach noted that "the formal rationale for construct validation sees a construct as defined by a network of relations, all of which are anchored to observables and so are testable."

Two problems present themselves for the validation process. One is the extent to which the attributes are and can be defined and the other is the adaptability of the measuring instruments. In order to address these problems as they relate to the evaluation of the Follow Through Program, my comments are grouped in the following four areas:

1. Political Ramifications
2. Insufficient Knowledge
3. Experimental Options
4. Instability of Sample

The controversy over whether or not the program should operate in a service or an experimental mode resulted in indecision on the part

of program developers and evaluators as to what information was required. Consequently, the evaluation assumed the role of trying to satisfy the needs of all interest groups. Politically, this was ill advised, since it became obvious that very few needs would be addressed by the evaluation for many reasons. It would not be possible to identify let alone construct and validate instruments in a multidimentional program like Follow Through. The learning domains and patterns of learning are still being identified and will continue for years to come. To many of the conceptualizers this is an expected event. For program evaluators it is a methodological nightmare. The planners only superficially considered the notion of testing for decision making, eventhough considerable effort was expended on testing. Had such consideration been made it is likely that the basic constructs would have been more carefully and narrowly defined and validated on the population of reference. It was not politically feasible to do the validation job needed in order to use test results in making decisions on the future course of the experimental program. This fact is now clear to many of us who have struggled with the problems of analysis and interpretations. Under the circumstances construct validation had to be assumed sufficient based on the standardized nature of the tests and the fact that the expected program impact would be great enough to overcome the marginal biases known to be in the tests. In the case of the Follow Through Experiment political sensitivity was more important than the issue of construct validity.

In the area of insufficient knowledge the early planners of the evaluation were not aware of the many constructs being addressed by program developers. For example one program developer pursued

transactional games theory to foster learning and improved self-image. Another pursued a self actualization concept as a precursor to learning. Some of the program developers realized the futility of testing for making policy decision and adopted a non-testing stance. Some program developers worked in a few domains and not in others. The various concepts evolved over time leaving the early evaluators with little information to design an acceptable test battery. However, given the nature of the program and the state of the art at that time, it is not likely that planners would have substantially improved on the test battery ultimately selected. There was too little knowledge about the particular learning problems of the populations of inquiry to accurately identify the constructs requiring measurements. Planners were obliged to start the evaluation before sufficient informantion was available. The lack of flexibility in choosing alternative test batteries, due primarily to the longitudinal nature of the evaluation, resulted in a test battery of questionable construct validity and ultimately inconclusive findings.

There were some twenty-two experimental options funded under the Follow Through Program. The replication of each of these options ranged ranged from one to 14. The decision for such inclusion was political. Even in 1969 no rational statistician would have recommended a planned variation experiment with haphazzard replications. Each option represented a different philosophy about teaching children from poor families. (See Table I for a list of Follow Through Models) An extensive battery of standardized tests was selected to try and measure the diverse goals and objectives presented by the 22 experimental options. (See Table II for a listing of the test battery) The gap was too wide between the perceived goals and the measurement instruments for the evaluation to

Table I

Follow Trough Models	No. Sites Implemented
APTAM Parent Implementation Education Model	8
Bank Street Model	14 *
Behavior Analysis Model	12 *
Behavior Oriented Prescriptive Training Approach	2
California Process Model	5 *
Cognitively Oriented Curriculum Model	10 *
Culturally Democratic Learning Environments Model	1
Cultural Linguistic Approach	3
Direct Instruction Model	15 *
EDC Open Education Follow Through Model	10 *
Florida Parent Education Model	11 *
Home-School Partnership Model	2
Individualized Early Learning Program	7 *
Interdependent Learning Model	2 *
Language Development Approach	6 *
Learning Systems Corporation Early Childhood Model	1
Mathemagenic Activities Program	6 *
New School Approach	4
Non-Graded Follow Through Model	4
Responsive Education Model	10 *
Tucson Early Education Model	19 *
Role-Trade Model	1

* Summarized in the national evaluation prepared by Abt Associates, Inc.

Table II

INSTRUMENTS USED IN THE NATIONAL EVALUATION*

When Administered	Type of Measure		
	Basic Skills	Cognitive-Conceptual Skills	Affective Skills
Entering Kindergarten or Entering First	Wide Range Achievement Test (WRAT) Peabody Picture Vocabulary Test (PPVT) Caldwell Pre-school Inventory (PSI)		
Leaving Kindergarten**	PPVT Metropolitan Test (MAT) - Primer . Listening for Sounds	MAT - Primer . Reading . Mathematics Concepts	Intellectual Achievement Responsibility Scale (IARS) . Locus of Control-positive . Locus of Control-negative Absence from School
Leaving First Grade**	MAT - Primary I . Word Knowledge . Word Analysis . Mathematics Computation	MAT - Primary I . Reading . Mathematics Concepts	Absence from School
Leaving Second Grade**	MAT - Primary II . Word Knowledge . Spelling . Mathematics Computation	MAT - Primary II . Reading . Mathematics Concepts . Mathematics Problem Solving	
Leaving Third Grade	MAT - Elementary . Word Knowledge . Spelling . Language (Part A & B) . Mathematics Computation	MAT - Elementary . Reading . Mathematics Concepts . Mathematics Problem Solving Raven's Coloured Progressive Matrics (Raven's)	IARS . Locus of Control-positive . Locus of Control-negative Coopersmith Self-Esteem Inventory (Coopersmith)

* Table includes only measures used in evaluation. For example, the WRAT was also administered at the end of kindergarten

** Measures used in evaluation of Cohort III K only.

produce valid results in the learning, psychological, and social domains. Measurements in the health and nutritional domains were, mercifully, not attempted.

The fourth area of commentary is instability of sample. This topic is identified as being a threat to construct validity because of its implications for making population estimates from samples. In the Follow Through Experiment attrition of sample exceeded 50% in many sites. Consequently, the population that started the program may have been quite different from the one that finished the third grade. The characteristics of the tests relative to the changing population is a critical area for more careful study and should be periodically examined over the course of a longitudinal study.

In summary the test battery used in the Follow Through Planned Variation Experiment had many threats to construct validity. The requirement to initiate an experiment before the experimental options were developed resulted in insufficient knowledge so vital to planning an evaluation design or developing and validating reliable test instruments. In social research this phenomena may be a way of life, since funds are allocate with immediate results expected. If test results are to be used to make major policy decisions, more attention must be given to construct validity. This may mean that a few experimental options be tried with rather narrowly defined constructs to focus the the evaluation. What we have noted over the years is that as theresults of our measurement programs fall short of expectation, the explications of caveats replace the discussion of findings resulting in a loss of crediability in the evaluation. Evaluation as an enterprise in education can not continue to survive this phenomena. We must produce valid reliable results.

REFERENCES

1. Cronbach, Lee J., "Validation of Educational Measures", Proceedings of the 1969 Invitational Conference on Testing Problems, Toward a Theory of Achievement Measurement, ETS, 1969, pp 35-52.

2. Van Dalen, Deolold B., Understanding Educational Research, Mc Graw Hill Book Co. 1973.

3. Wholey, Joseph, et. al., "Federal Evaluation Policy" The Urban Institute, Washington, DC, 1970.

4. Rivlin, A. and Timpane, P. M., "Planned Variation in Education", The Brookings Institution, Washington, DC, 1975.